NEOCARE
HEALTH IN A GLANCE

NEOCARE

RPPG BASED NEONATAL HEALTH MONITORING SYSTEM

Members

Lasitha Amarasinghe
Sahan Abeyrathna
Induwara Morawakgoda
Yasiru Alahakoon

Supervisors

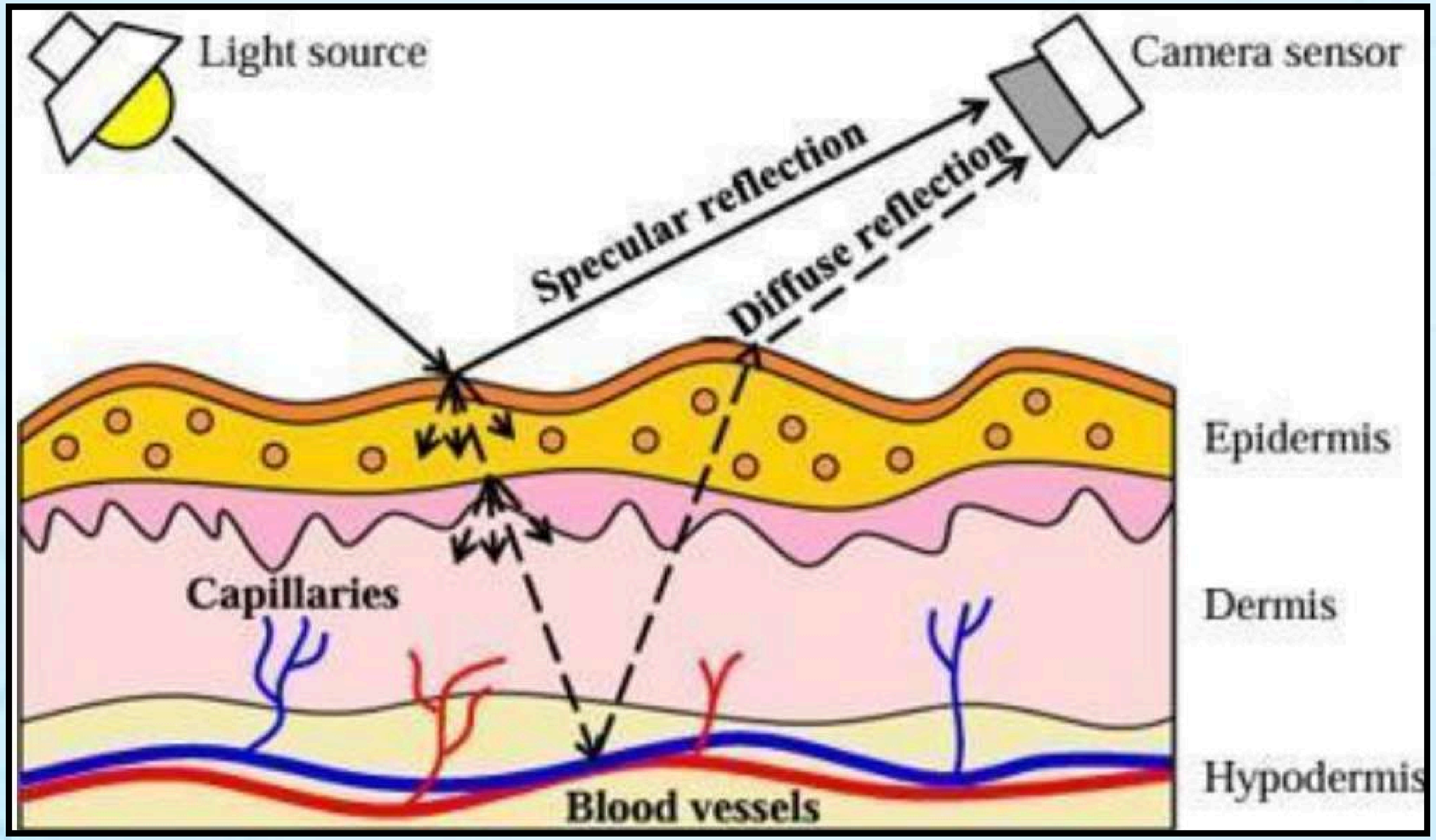
Dr. Sampath K.Perera
Dr. Pranjeevan Kulasingham
Prof. Anusha WIthana (University of Sydney)
Prof. Nishani Lucas (Consultant Neonatologist, De Soysa Hospital for Women, Colombo)

ENTC
Electronic and Telecommunication Engineering

THE UNIVERSITY OF SYDNEY

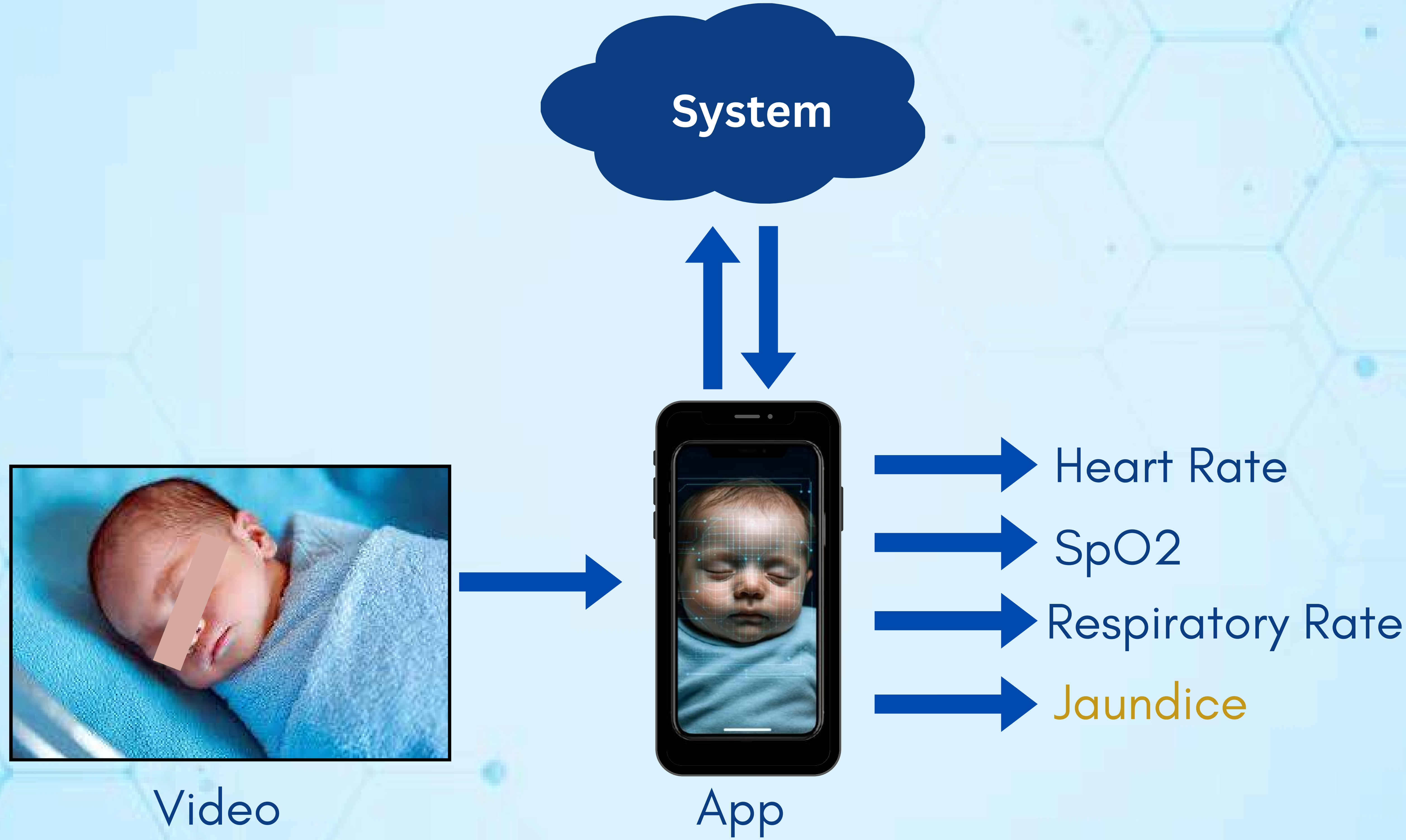
What is RPPG ?

Remote Photoplethysmography is a contactless method to measure vital signs by tracking skin color changes from blood flow.



Wang, W., den Brinker, A. C., Stuijk, S., & de Haan, G. (2017). Algorithmic Principles of Remote PPG. IEEE Transactions on Biomedical Engineering, 64(7). https://doi.org/10.1109/TBME.2016.2609282

PROPOSED APPROACH



System

Video

App

Heart Rate

SpO2

Respiratory Rate

Jaundice

https://www.youtube.com/watch?v=_6Log84TbNQ&list=PPSV

JAUNDICE

- Excess bilirubin in the blood
- Newborns and premature babies
- Yellowing of skin and eyes
- Brain damage (kernicterus) if untreated

https://www.youtube.com/watch?v=_6Log84TbNQ&list=PPSVhttps://www.nidirect.gov.uk/conditions/newborn-jaundice

PROBLEMS

Inaccurate chest count and manually counting breaths introduces human error.

https://www.youtube.com/watch?v=_6Log84TbNQ&list=PPSV

Touch-based devices cause harm and can spread infection and irritate delicate skin.

L. Zhou, M. Guess, K. R. Kim, and W.-H. Yeo, "Skin-interfacing wearable biosensors for smart health monitoring of infants and neonates," Communications Materials, vol. 5, no. 1, pp. 1-13, May 2024, doi: https://doi.org/10.1038/s43246-024-00511-6.
"Medical Adhesive-Related Skin Injuries (MARSI)," WoundSource, Jun. 02, 2017. https://www.woundsource.com/patientcondition/medical-adhesive-related-skin-injuries-marsi

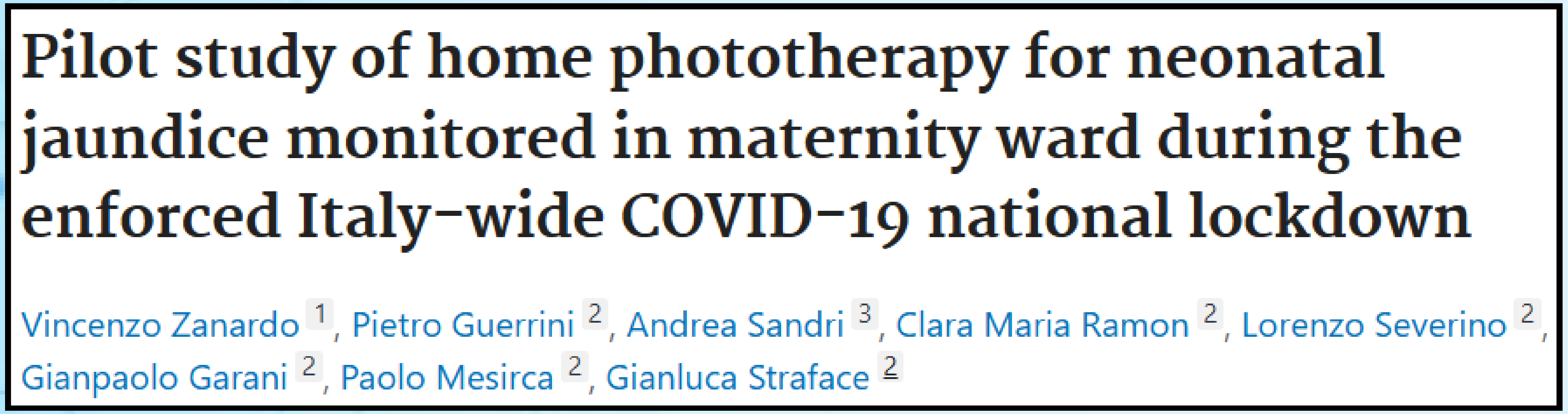
Physical contact is risky for premature or disabled babies in incubators, making vital sign monitoring difficult.

L. Zhou, M. Guess, K. R. Kim, and W.-H. Yeo, “Skin-interfacing wearable biosensors for smart health monitoring of infants and neonates,” Communications Materials, vol. 5, no. 1, pp. 1-13, May 2024, doi: https://doi.org/10.1038/s43246-024-00511-6.
“Medical Adhesive-Related Skin Injuries (MARSI),” WoundSource, Jun. 02, 2017. https://www.woundsource.com/patientcondition/medical-adhesive-related-skin-injuries-marsi

Jaundice in newborns is usually detected by blood tests, which are invasive.

https://www.ausmed.com/learn/articles/neonatal-jaundice
https://en.wikipedia.org/wiki/Neonatal_jaundice
https://www.rch.org.au/clinicalguide/guideline_index/jaundice_in_early_infancy/

Parents lack a quick and reliable way to check their baby's health at home during emergencies.



Pilot study of home phototherapy for neonatal jaundice monitored in maternity ward during the enforced Italy-wide COVID-19 national lockdown

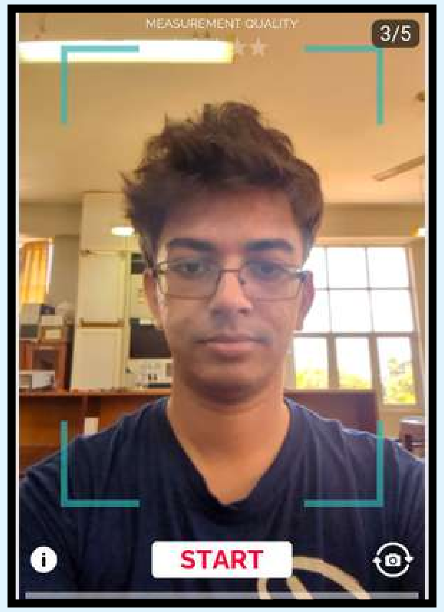
Vincenzo Zanardo[1], Pietro Guerrini[2], Andrea Sandri[3], Clara Maria Ramon[2], Lorenzo Severino[2], Gianpaolo Garani[2], Paolo Mesirca[2], Gianluca Straface[2]

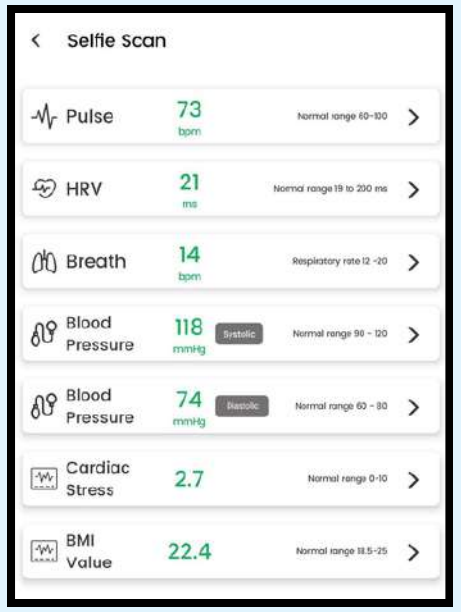
https://medpol.providers.amerigroup.com/medpolicies/amerigroup/active/gl_pw_a053626.html
https://www.thewomens.org.au/health-information/at-home-with-your-baby/phototherapy-at-home

EXISTING SOLUTIONS

Bilirubinometers

Philips Lumify





Selfie Scan

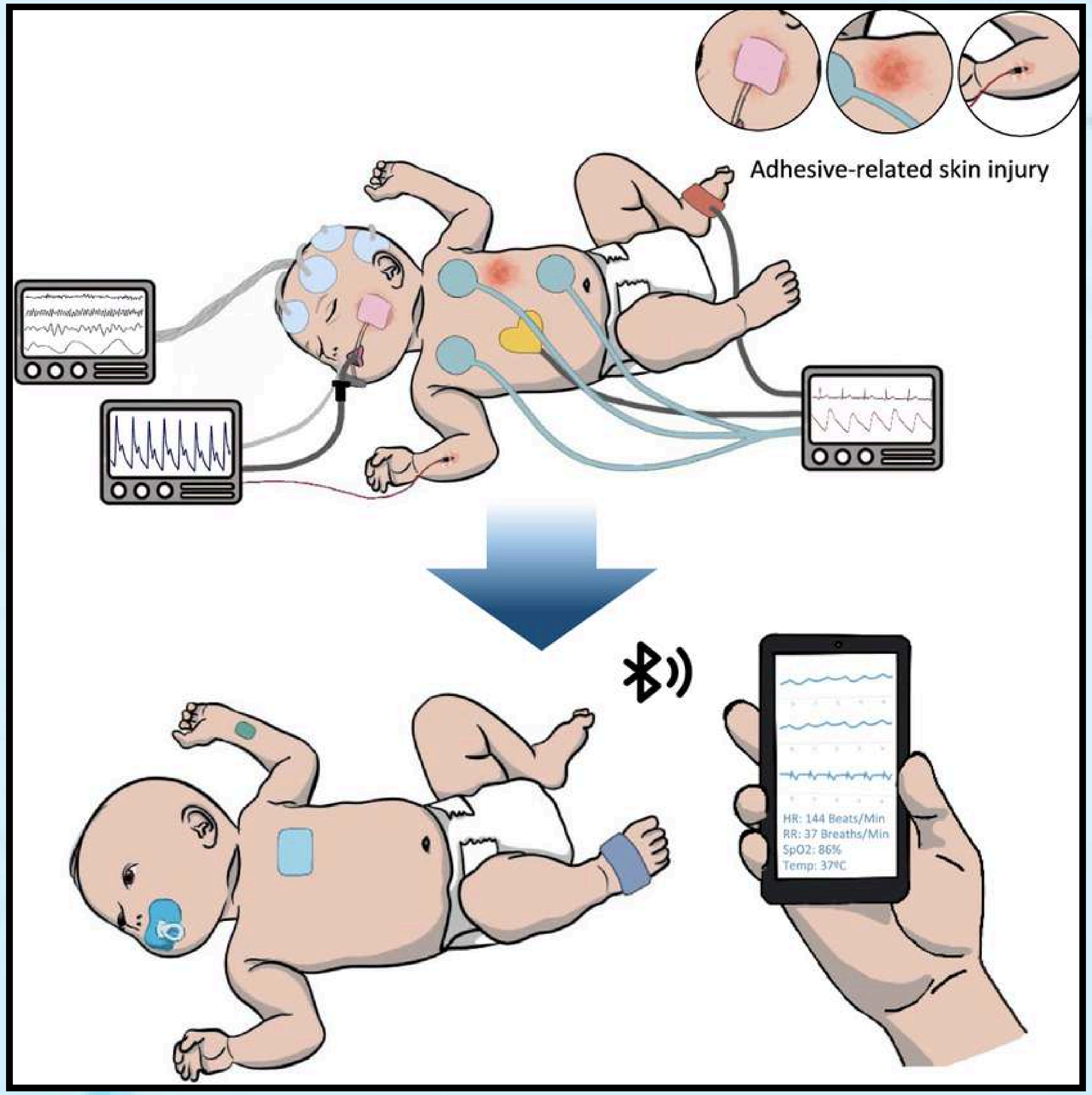
Pulse	73 bpm		Normal range 60-100	>
HRV	21 ms		Normal range 19 to 200 ms	>
Breath	14 bpm		Respiratory rate 12 -20	>
Blood Pressure	118 mmHg	Systolic	Normal range 90 - 120	>
Blood Pressure	74 mmHg	Diastolic	Normal range 60 - 80	>
Cardiac Stress	2.7		Normal range 0-10	>
BMI Value	22.4		Normal range 18.5-25	>

Doc990 Selfie Scan by Dialog

https://www.philips.co.uk/healthcare/resources/landing/lumifywithreacts
https://dialog.lk/news/dialogs-digital-health-launches-sri-lankas-first-ai-based-health-scan?language=ta
https://ubicomplab.cs.washington.edu/pdfs/bilicam.pdf

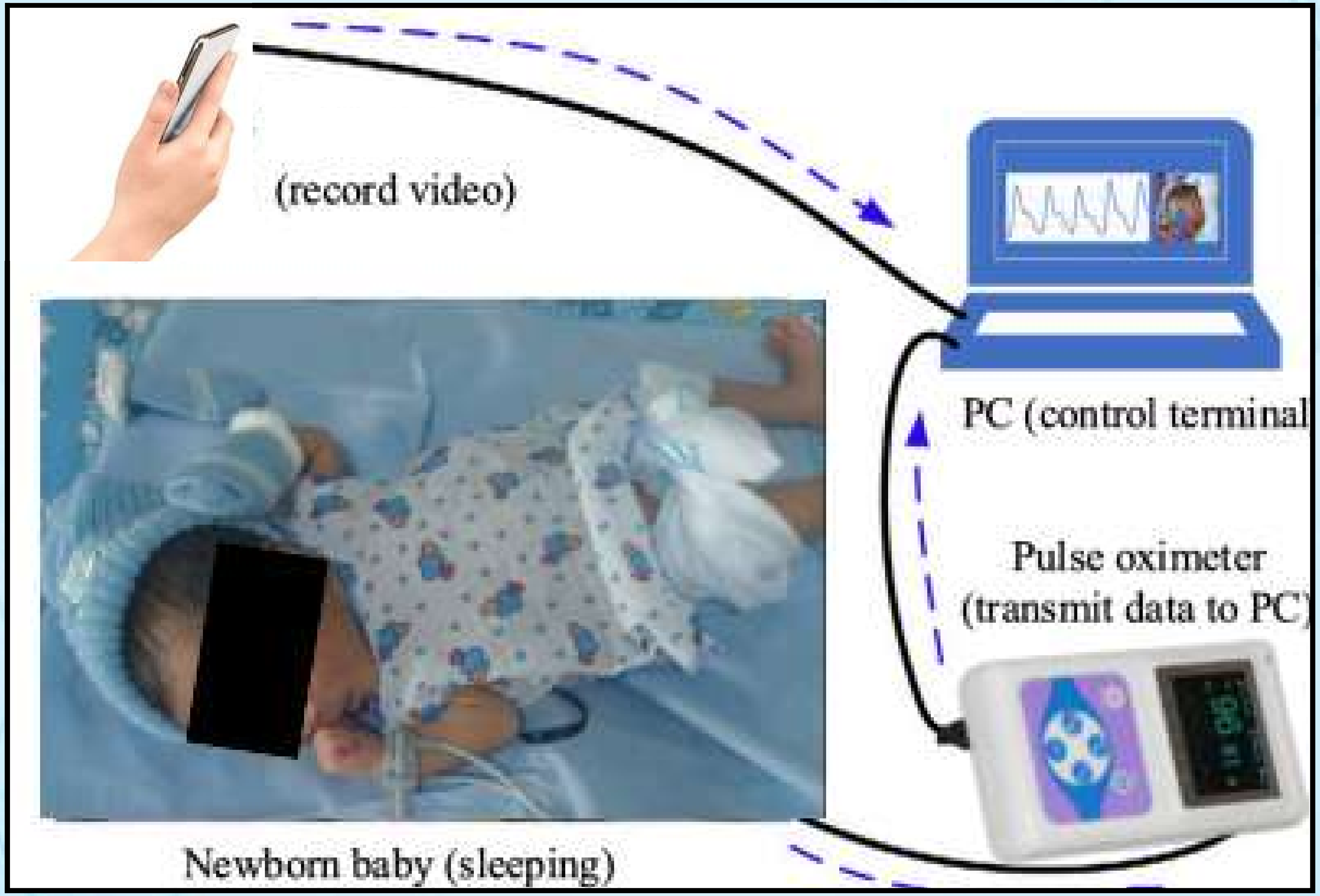
DELIVERABLES

Enhance existing non-contact, video-based algorithms to accurately estimate HR, SpO_2, RR, jaundice status in neonates



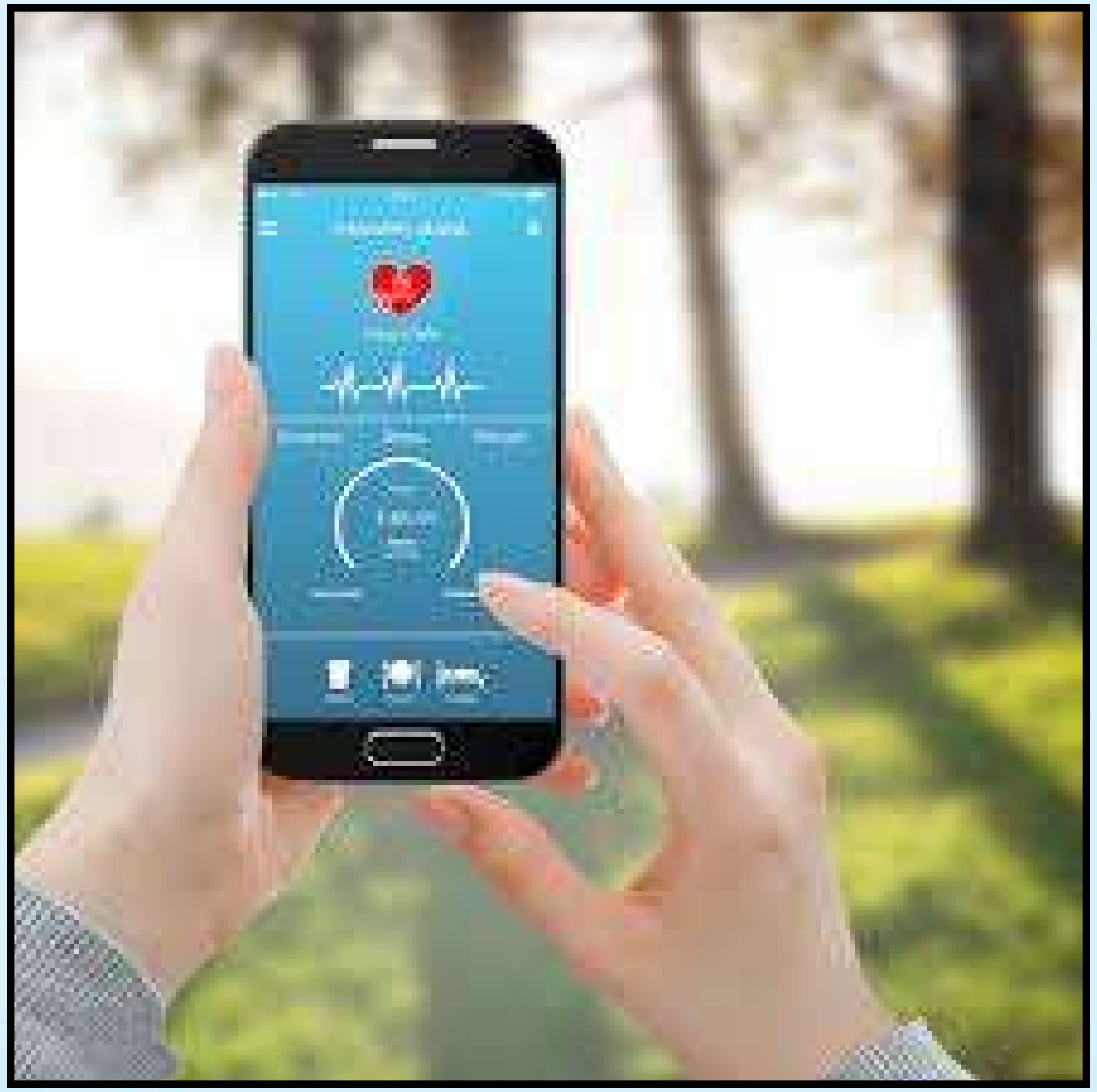
https://www.nature.com/articles/s43246-024-00511-6

Create a new dataset specifically for neonates to validate these algorithms


[redacted]

https://www.researchgate.net/figure/Principle-of-remote-photoplethysmography-rPPG-based-on-the-dichromatic-reflection-model_fig1_354758072?utm_source=chatgpt.com

Mobile application for contactless monitoring of neonates



https://nedis.co.uk/en-gb/nedis-smartlife-health

Implement privacy preserving techniques and secure data handling

	Pros	Cons
Differential Privacy	Lightweight for mobile use.	Can reduce accuracy of predictions if too much noise is added.
Federated Learning	Supports on device learning.	Requires coordination and bandwidth for training.
Homomorphic Encryption	Strong cryptographic guarantees.	Computationally intensive
Secure Multi-Party Computation	Strong privacy in collaborative computations.	Complex to implement.

RESEARCH GAPS

RESEARCH GAPS

1. Few researches on neonates
2. Skin tone bias for lighter skin colors
3. Low explainability
4. No system with multiple output including Spo2,HR, RR and Jaundice
5. Motion and segmentation
6. No models or datasets that are collected by mobile phones for neonatal rppg estimation
7. Lack of methods that preserve the privacy

PREVIOUS TEAM'S WORK

Previous Team

Neonatal Heart Rate Estimation on NBHR Dataset → **lowest MAE**
lowest MAPE

Neonatal SpO2 Level Estimation on NBHR Dataset → **lowest MAE**
lowest RMSE

Neonatal Facial Video Dataset → **VideoPulse**

NOVELTY

NOVELTY

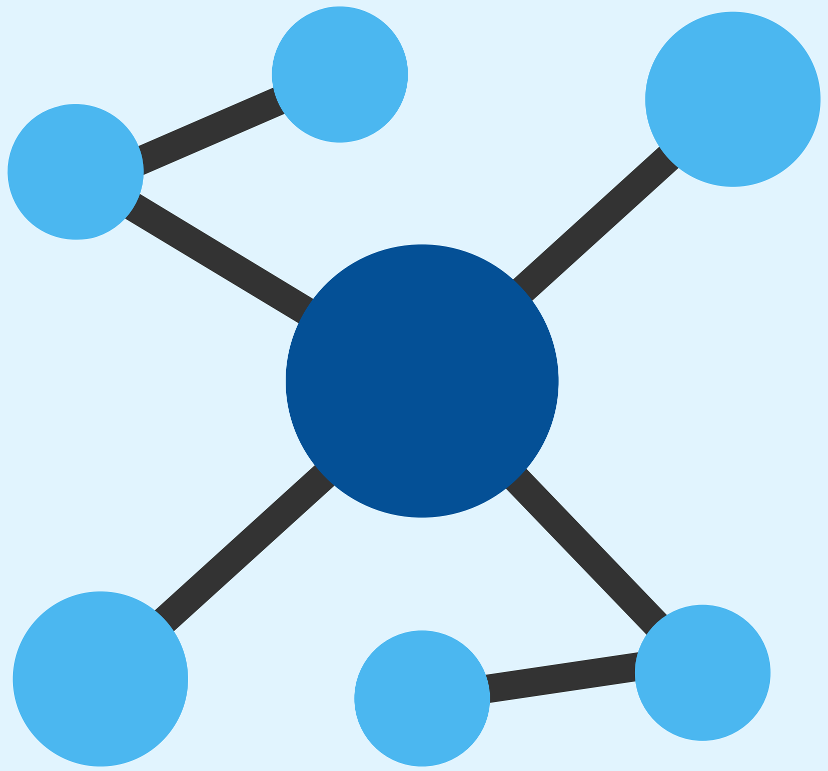
First method on neonatal respiratory rate prediction using non invasive methods

First mobile app to embed several vital signs including RR, HR, SpO2 level for neonates

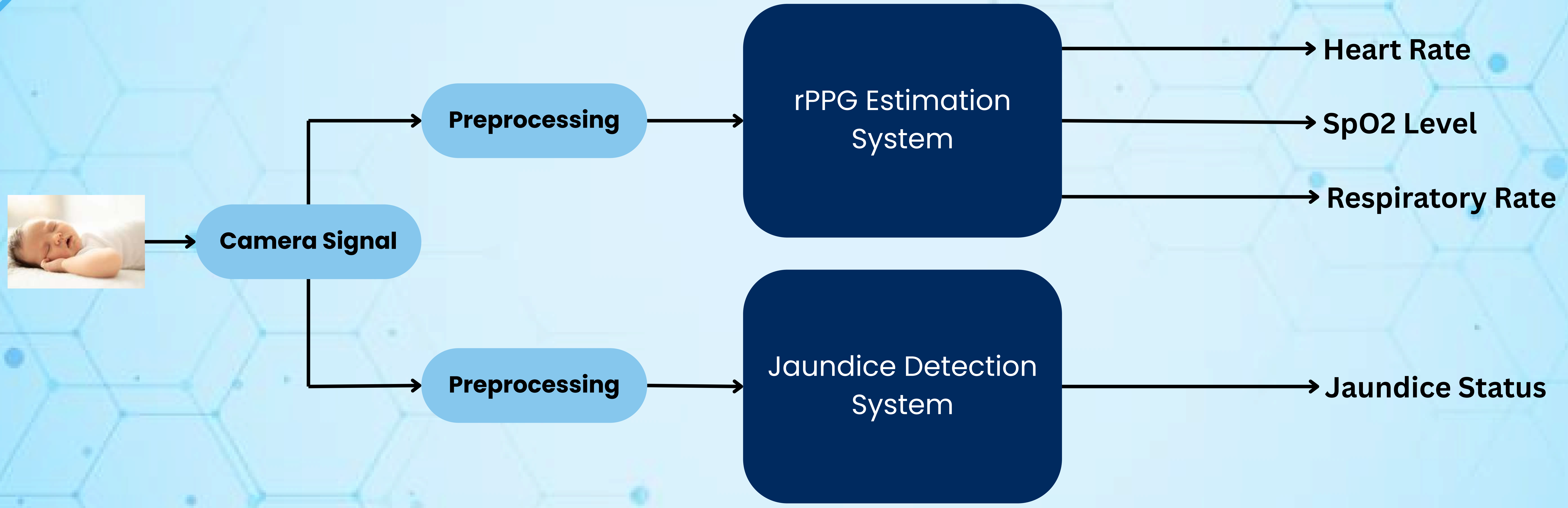
Creating a new dataset on neonates that supports HR, SpO2 and RR

X. Liu, G. Narayanswamy, A. Paruchuri, X. Zhang, J. Tang, Y. Zhang, R. Sengupta, S. Patel, Y. Wang, and D. McDuff, “rppg-toolbox: Deep remote PPG toolbox,” Advances in Neural Information Processing Systems, vol. 36, pp. 68485–68510, 2023.

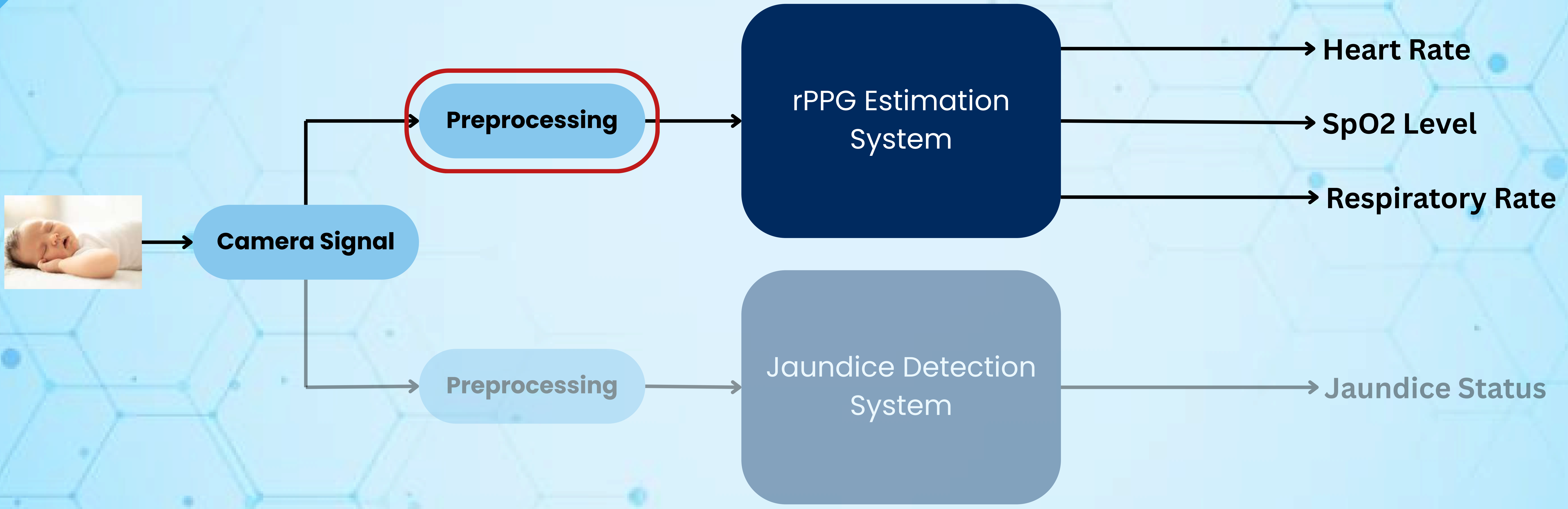
ALGORITHM DEVELOPMENT



Algorithm Development



Algorithm Development



Camera Signal

Preprocessing

rPPG Estimation System

Heart Rate

SpO2 Level

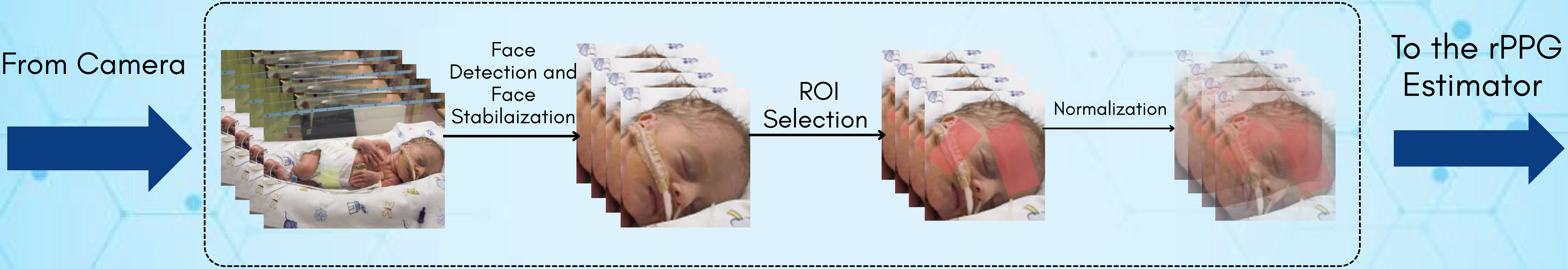
Respiratory Rate

Preprocessing

Jaundice Detection System

Jaundice Status

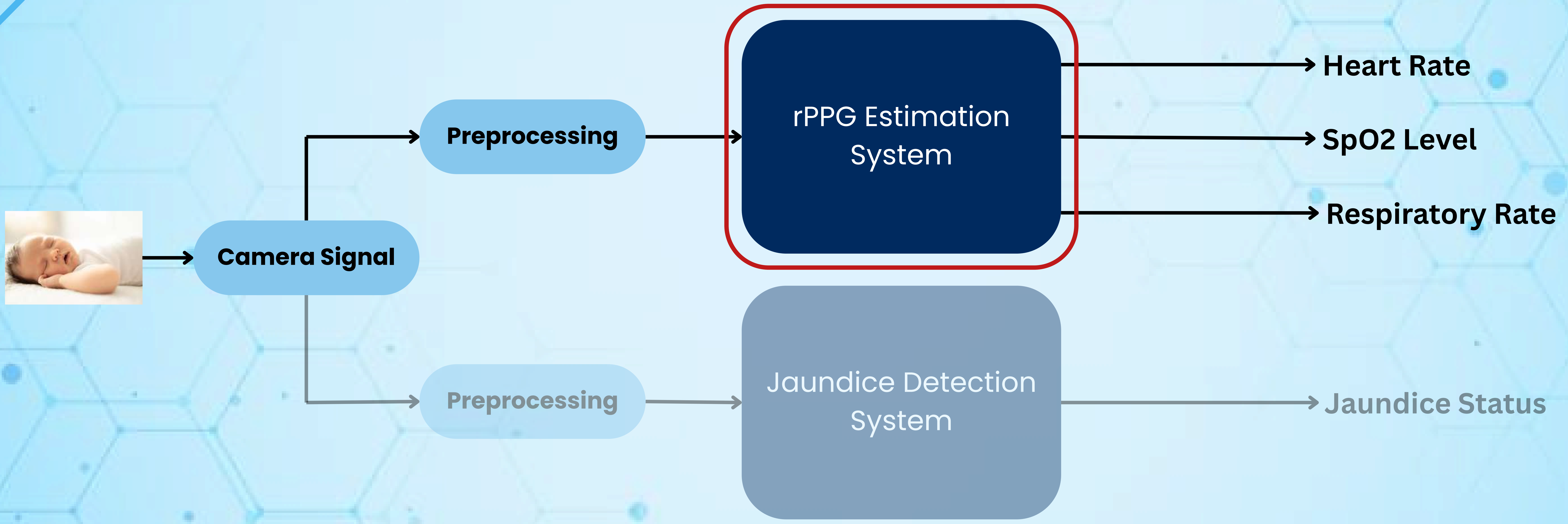
Preprocessing - rPPG



- **Region of Interest(ROI) : Forehead and Cheeks**

Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. Optics express, 2008.
X. Liu, G. Narayanswamy, A. Paruchuri, X. Zhang, J. Tang, Y. Zhang, R. Sengupta, S. Patel, Y. Wang, and D. McDuff, “rppg-toolbox: Deep remote PPG toolbox,” Advances in Neural Information Processing Systems, vol. 36, pp. 68485–68510, 2023.
C. Hu, K.-Y. Zhang, T. Yao, S. Ding, J. Li, F. Huang, and L. Ma, “An end-to-end efficient framework for remote physiological signal sensing,” in 2021 IEEE/CVF In ternational Conference on Computer Vision Workshops (ICCVW), 2021, pp. 2378 2384.

Algorithm Development



Camera Signal

Preprocessing

rPPG Estimation System

Heart Rate

SpO2 Level

Respiratory Rate

Preprocessing

Jaundice Detection System

Jaundice Status

Comparison of rPPG Estimation Methods

	Signal Processing Approaches	Deep Learning Based Approaches
Data Used	Color signals	Video frames
Noise Handling	Rule-based filtering	Learns from data
Motion Robustness	Sensitive	Robust
Adaptability	Limited	Adaptive
Accuracy	Low	High

X. Liu, G. Narayanswamy, A. Paruchuri, X. Zhang, J. Tang, Y. Zhang, R. Sengupta, S. Patel, Y. Wang, and D. McDuff, “rppg-toolbox: Deep remote PPG toolbox,” Advances in Neural Information Processing Systems, vol. 36, pp. 68485–68510, 2023.

W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, “Algorithmic principles of remote ppg,” IEEE Transactions on Biomedical Engineering, vol. 64, no. 7, pp. 1479–1491, 2017.

Comparison of Deep Learning based Architectures for rPPG Estimation

	3D CNN	Transformers(ViT)	Mamba
Compute Cost	Low	High(quadratic)	Low
Sequence Handling	Short-Medium	Long	Long
Accuracy	Moderate	High	High

X. Liu, G. Narayanswamy, A. Paruchuri, X. Zhang, J. Tang, Y. Zhang, R. Sengupta, S. Patel, Y. Wang, and D. McDuff, “rppg-toolbox: Deep remote PPG toolbox,” Advances in Neural Information Processing Systems, vol. 36, pp. 68485–68510, 2023.
A. Gu and T. Dao, “Mamba: Linear-time sequence modeling with selective state spaces,” 2024. [Online]. Available: https://openreview.net/forum?id=AL1fq05o7H
C. Hu, K.-Y. Zhang, T. Yao, S. Ding, J. Li, F. Huang, and L. Ma, “An end-to-end efficient framework for remote physiological signal sensing,” in 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021, pp. 2378 2384.

Heart Rate Estimation Methods

Method	Architecture	MAE(bpm) on PURE Dataset	Parameter Size(M)
PhysMamba	Mamba	**0.25**	0.56
RhythmFormer	ViT	0.27	3.25
FactorizePhys	3D CNN	1.04	**0.05**

For Adults

Method	Architecture	MAE(bpm) on NBHR Dataset	Parameter Size(M)
NBHRnet-2s	2D CNN + LSTM	3.97	0.91
PhysNet(by Previous Group)	3D CNN	**2.97**	**0.77**

For Neonates

X. Liu, G. Narayanswamy, A. Paruchuri, X. Zhang, J. Tang, Y. Zhang, R. Sengupta, S. Patel, Y. Wang, and D. McDuff, “rppg-toolbox: Deep remote PPG toolbox,” Advances in Neural Information Processing Systems, vol. 36, pp. 68485–68510, 2023.

B. Huang, W. Chen, C.-L. Lin, C.-F. Juang, Y. Xing, Y. Wang, and J. Wang, “A neonatal dataset and benchmark for non-contact neonatal heart rate monitoring based on spatio-temporal neural networks,” Engineering Applications of Artificial Intelligence, vol. 106, p. 104447, 2021, doi: 10.1016/j.engappai.2021.104447.

Respiratory Rate Estimation Methods

method	Architecture	RMSE(Hz) on OBF Dataset	Parameter Size(M)
CVD	2D CNN + 1D Conv	0.058	8.53
rPPGNet	3D CNN	0.064	**0.45**
PhysFormer	ViT	**0.054**	7.03

For Adults

- **Currently, there are no established methods for estimating the respiratory rate in neonates using the rPPG signal**

Xuesong Niu, Zitong Yu, Hu Han, Xiaobai Li, Shiguang Shan, and Guoying Zhao. Video-based remote physiological measurement via cross-verified feature disentangling. In ECCV, pages 295–310. Springer, 2020.
Zitong Yu, Wei Peng, Xiaobai Li, Xiaopeng Hong, and Guoying Zhao. Remote heart rate measurement from highly compressed facial videos: an end-to-end deep learning solution with video enhancement. In ICCV, 2019.
Z. Yu, Y. Shen, J. Shi, H. Zhao, P. H. S. Torr, and G. Zhao, "PhysFormer: Facial video-based physiological measurement with temporal difference transformer," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 4186–4196.

SPO2 Estimation Methods

Method	MAE (%) on VIPL-HR Dataset
Past Analytic (Ratio of Ratios)	3.334
EfficientNet-B3 + RGB	1.274
Multi Model Fusion Method	**1.000**

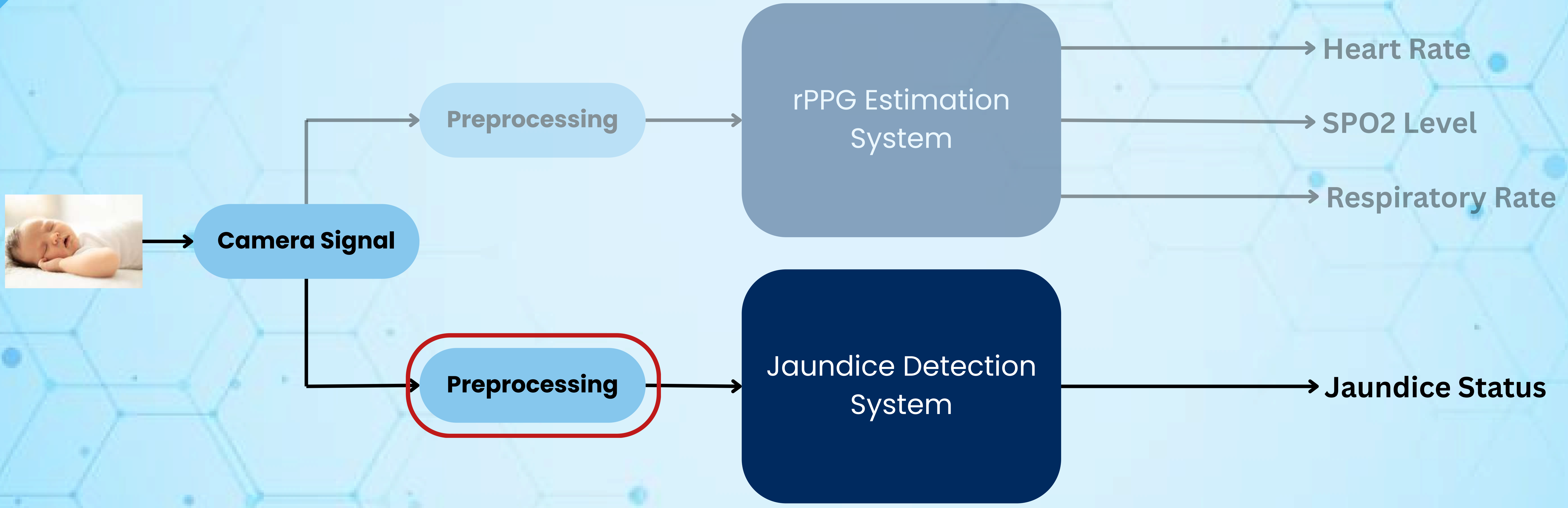
For Adults

Method	MAE (%) on NBHR Dataset
Modified Physnet(Previous Group)	**1.69**

For Neonates

C.-H. Cheng, Z. Yuen, S. Chen, K.-L. Wong, J.-W. Chin, T.-T. Chan, and R. So, “Contactless blood oxygen saturation estimation from facial videos using deep learning,” Bioengineering, vol. 11, no. 3, p. 251, 2024.

Algorithm Development



Camera Signal

Preprocessing

rPPG Estimation System

Heart Rate

SPO2 Level

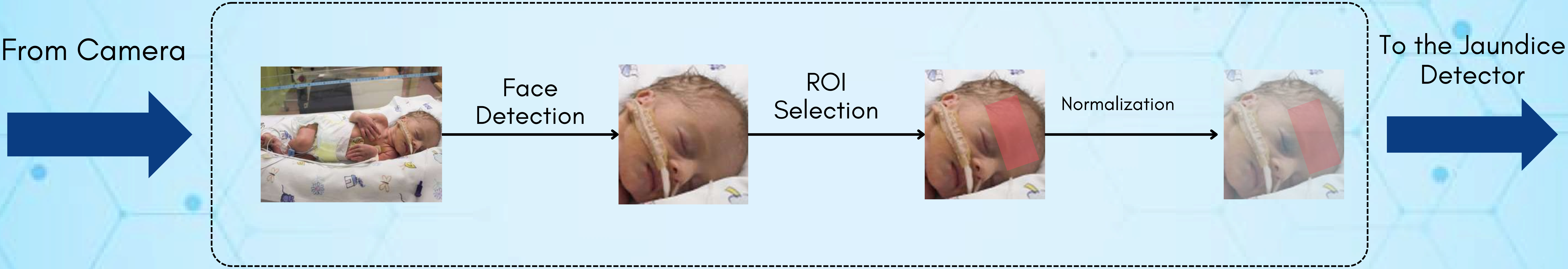
Respiratory Rate

Preprocessing

Jaundice Detection System

Jaundice Status

Preprocessing - Jaundice



From Camera

Face Detection

ROI Selection

Normalization

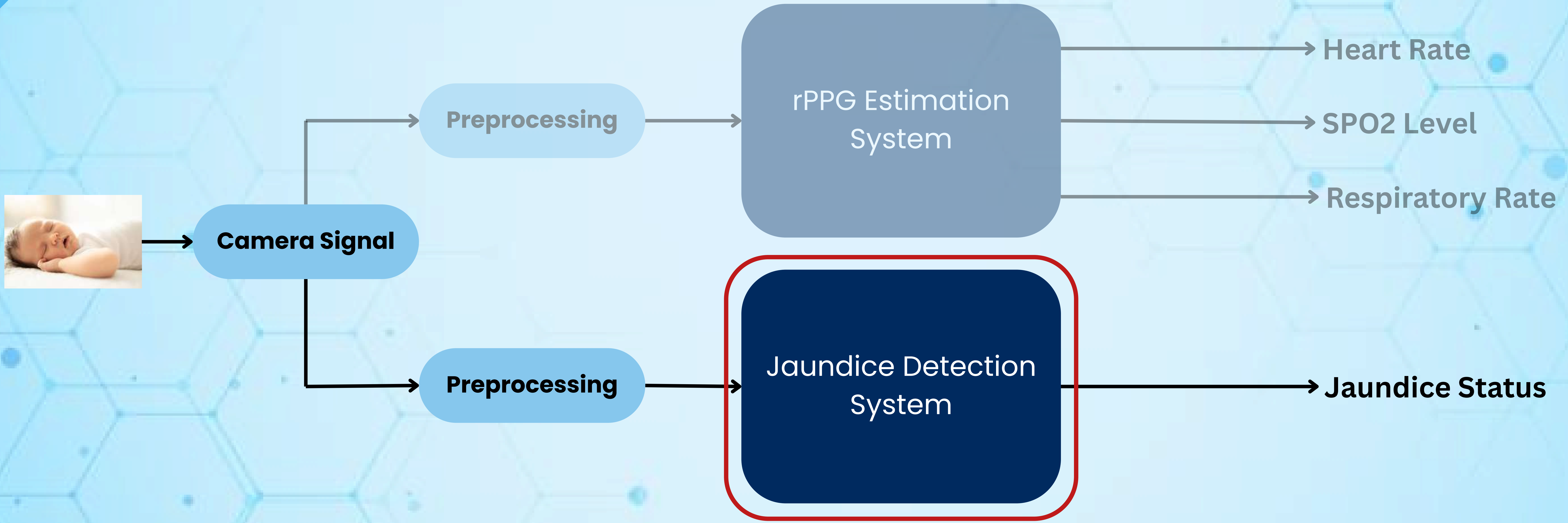
To the Jaundice Detector

- **Region of Interest(ROI) : Forehead**

Khanam F-T-Z, Al-Naji A, Perera AG, Wang D, Chahl J. Non-invasive and non-contact automatic jaundice detection of infants based on random forest. Comput Methods Biomech Biomed Eng: Imagin

Salami FO, Muzammel M, Mourchid Y, Othmani A. Artificial Intelligence non-invasive methods for neonatal jaundice detection: A review. Artif Intell Med. 2025 Apr;162:103088. doi: 10.1016/j.artmed.2025.103088. Epub 2025 Feb 19. PMID: 39988547.

Algorithm Development



Camera Signal

Preprocessing

rPPG Estimation System

Heart Rate

SPO2 Level

Respiratory Rate

Preprocessing

Jaundice Detection System

Jaundice Status

Comparison of Non-Invasive Approaches for Neonatal Jaundice Detection

	Machine Learning (ML)	Deep Learning (DL)
Feature Handling	Manual feature selection	Learns features from images
Complexity	Low	High
Adaptability	Limited	Adaptive
Accuracy	Low	High

Salami FO, Muzammel M, Mourchid Y, Othmani A. Artificial Intelligence non-invasive methods for neonatal jaundice detection: A review. Artif Intell Med. 2025 Apr;162:103088. doi: 10.1016/j.artmed.2025.103088. Epub 2025 Feb 19. PMID: 39988547.

Deep Learning-Based Studies for Neonatal Jaundice Detection

Study	**Accuracy**
ResNet50 with image augmentation	84.1%
Smartphone-based skin color segmentation	93.0%
Spectral-Spatial Graph Neural Network (SSGNN)	**96.5%**

Salami FO, Muzammel M, Mourchid Y, Othmani A. Artificial Intelligence non-invasive methods for neonatal jaundice detection: A review. Artif Intell Med. 2025 Apr;162:103088. doi: 10.1016/j.artmed.2025.103088. Epub 2025 Feb 19. PMID: 39988547.

DATA COLLECTION

Available Datasets

Neonatal rPPG Estimation

	NBHR	Videopulse
No of Neonates	257	52
Total Video Length (hours)	9.6	2.6
No of Videos	1130	157
Information	Synchronized vital signs (PPG, Heart Rate, SpO_2)	2–3 recording scenarios per neonate

Huang, Bin & Chen, Weihai & Lin, Chun-Liang & Juang, Chia-Feng & Xing, Yuanping & Wang, Yanting & Wang, Jianhua. (2021). A neonatal dataset and benchmark for non-contact neonatal heart rate monitoring based on spatio-temporal neural networks. Engineering Applications of Artificial Intelligence. 106. 104447. 10.1016/j.engappai.2021.104447.

Available Datasets

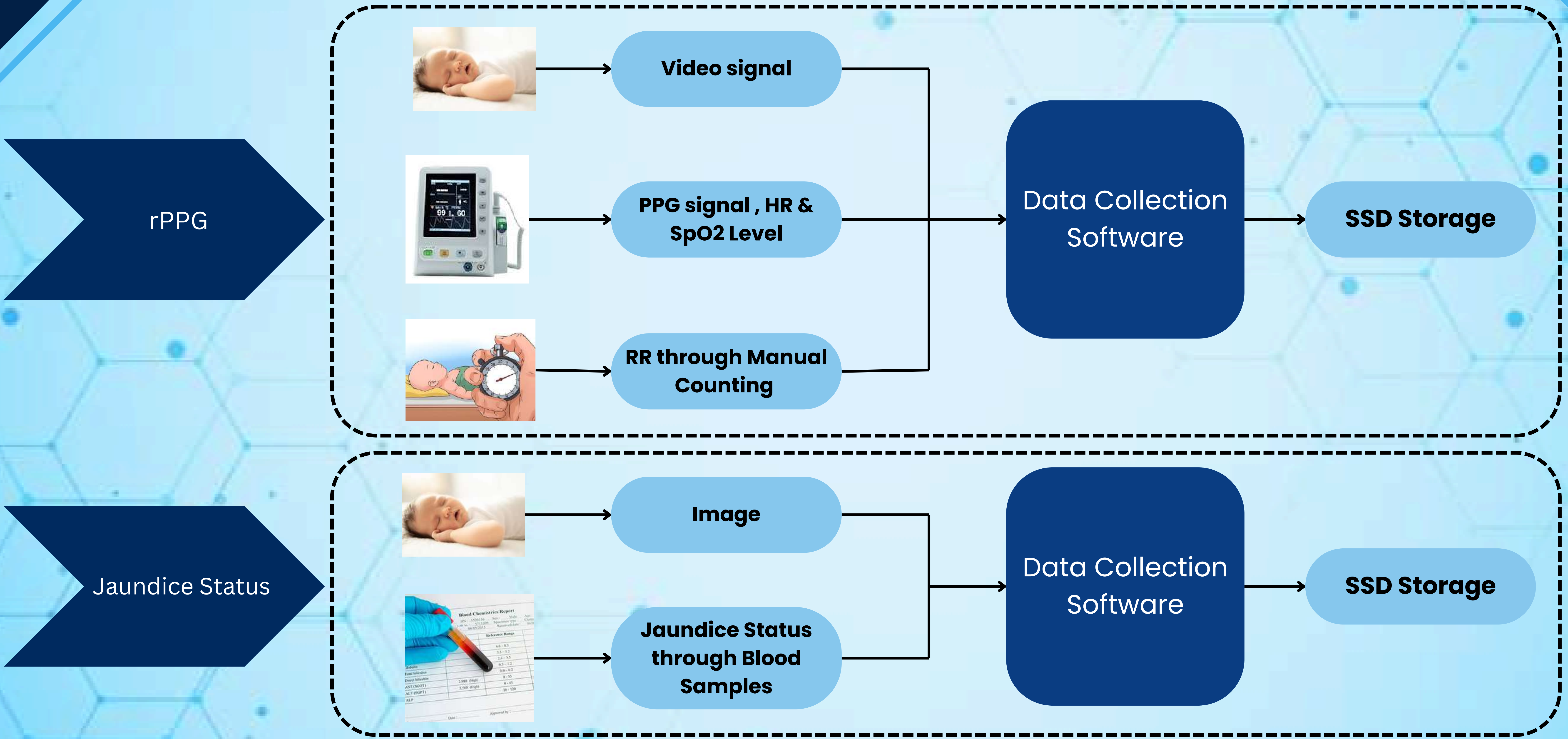
Neonatal Jaundice Detection

	NJN	NJA
No of Images	760 (560 normal, 200 jaundiced)	600 (forehead & sternum)
Resolution (pixels)	1000 × 1000	640 x 480
Collected device	iPhone 11 Pro Max	digital microscope connected to a smartphone
Collected place	NICU, Iraq	Ghaem Hospital, Iran

Abdulrazzak, A. Y., Mohammed, S. L., & Al-Naji, A. (2023). NJN: A Dataset for the Normal and Jaundiced Newborns. BioMedInformatics, 3(3), 543–552. https://doi.org/10.3390/biomedinformatics3030037

Makhloughi, F., Sadeghi Bajestani, G., & Faramarzi, R. (2024). Neonatal Jaundice Assessment: A Dataset of Forehead and Sternum Images for Bilirubin Estimation (Version 1) [Dataset]. Mendeley Data. https://doi.org/10.17632/yfsz6c36vc.1

Data Collection Process



rPPG

Video signal

PPG signal , HR & SpO2 Level

RR through Manual Counting

Data Collection Software

SSD Storage

Jaundice Status

Image

Jaundice Status through Blood Samples

Data Collection Software

SSD Storage

Data Collection Details

Description	
Total number of neonates	100
	4
Total number of recording sessions	400
Video length per session	1 minute

- Data collection will be conducted under the supervision of Consultant Neonatologist Dr. Nishani Lucus, with the assistance of professional medical staff

De Soyza Hospital Environment (Post Natal Ward)

Full Unit Bed

Neonate Bed

PLAN & PROGRESS

Ethical Clearnace

University

Hospital

University Ethics Review Committee

De Soysa Hospital for Women Ethics Review Committee

- Application for Ethical Clearance ✓
- Project Description ✓
- Information Sheet ✓
- Consent Forms ✓

- ERC Application
- ERC Application Checklist
- Research Proposal

Budget

Item	Price
Samsung S25 Ultra Phone	USD 876 (Rs. 264413.59)
1TB portable SSD Storage	USD 96 (Rs. 28976.83)
Other Expences	USD 70 (Rs. 21128.94)
Total	**USD 1042 (Rs. 313382.80)**

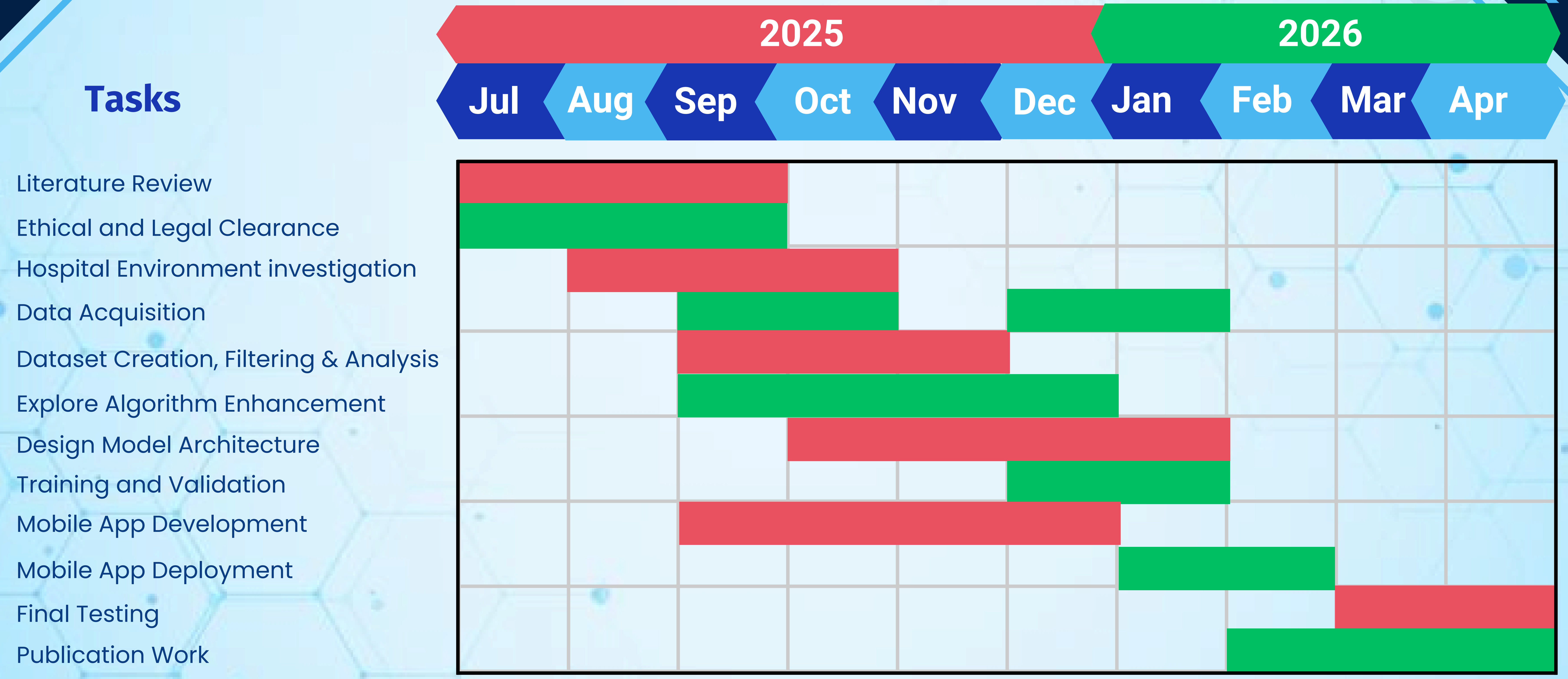
- The GPU resources are provided by our co-supervisor Dr. Anusha Withana from University of Sydney.
- Other expenses include documentation & transportation fees.

Task Allocation

🟥 Main 🟩 Sub

Task	Lasitha	Sahan	Induwara	Yasiru
Literature Review, Documentation	🟩	🟩	🟩	🟩
Neonatal Data Aquisition	🟩	🟩	🟩	🟩
Noise Analysis & Filter Design		🟩		🟥
Data Filtering & Analysis	🟥		🟩	
Data pre-processing & Feature Extraction			🟩	🟥
Jaundice Detection Model Development	🟥	🟩		
Rasperatory Rate Prediction Algorithm Enhancement			🟥	🟩
Heart Rate & SPO2 Level Predction Algorithm Enhancement		🟥		🟩
Model Pipeline Design	🟩		🟥	
Mobile App Development & Deployment	🟩	🟥		

Time Line



Tasks	2025						2026			
	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Literature Review	🟥	🟥	🟥							
Ethical and Legal Clearance	🟩	🟩	🟩							
Hospital Environment investigation		🟥	🟥	🟥						
Data Acquisition			🟩	🟩		🟩	🟩			
Dataset Creation, Filtering & Analysis			🟥	🟥	🟥					
Explore Algorithm Enhancement			🟩	🟩	🟩	🟩				
Design Model Architecture				🟥	🟥	🟥	🟥			
Training and Validation						🟩	🟩			
Mobile App Development			🟥	🟥	🟥	🟥				
Mobile App Deployment							🟩	🟩		
Final Testing									🟥	🟥
Publication Work								🟩	🟩	🟩

Risk Factors

Risk Type	Potential Risks	How to Mitigate?
Operational	Delays in ethical clearance process for data collection	Started the ethical clearance process early
Technical	Motion artifacts, lighting variations	Stabilization methods, normalization
Ethical	Sensitive data handling, data privacy	Getting approval from each mother, Privacy preserving techniques

THANK YOU

DISCUSSION

SUPPORT SLIDES

PREVIOUS TEAM

Comparison of heart rate estimation methods on NHBR dataset

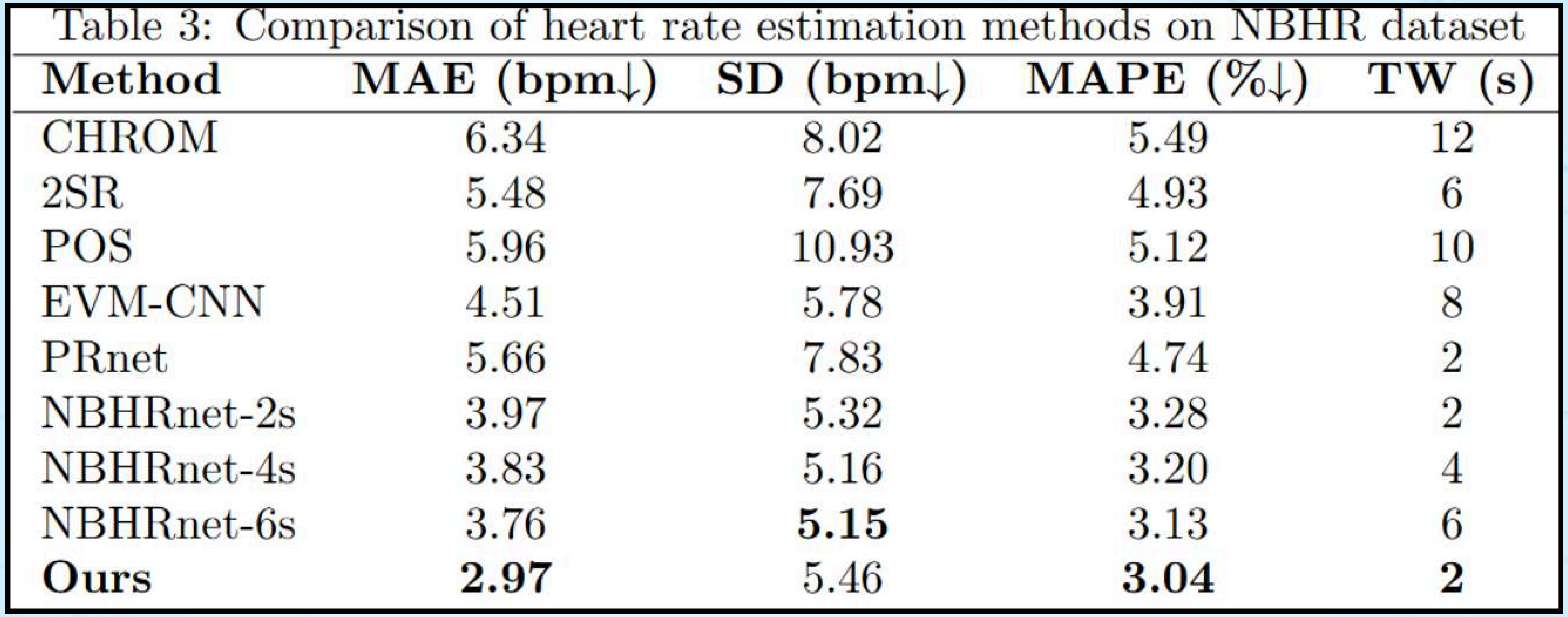
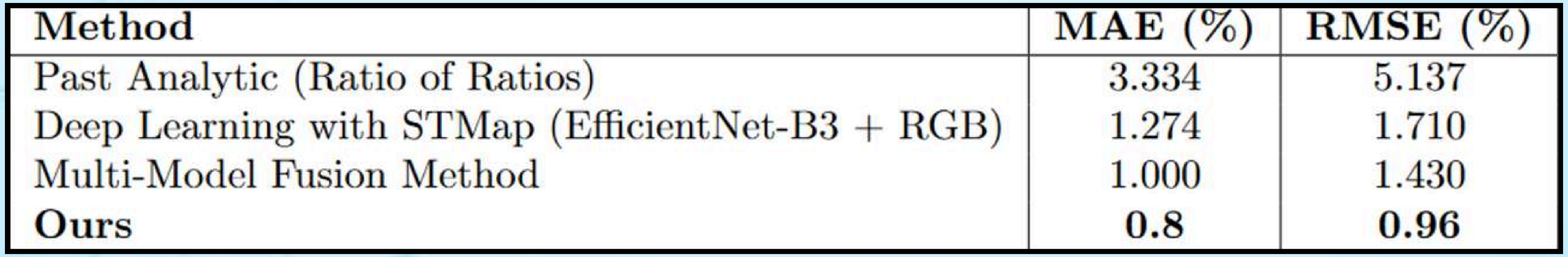


Table 3: Comparison of heart rate estimation methods on NBHR dataset

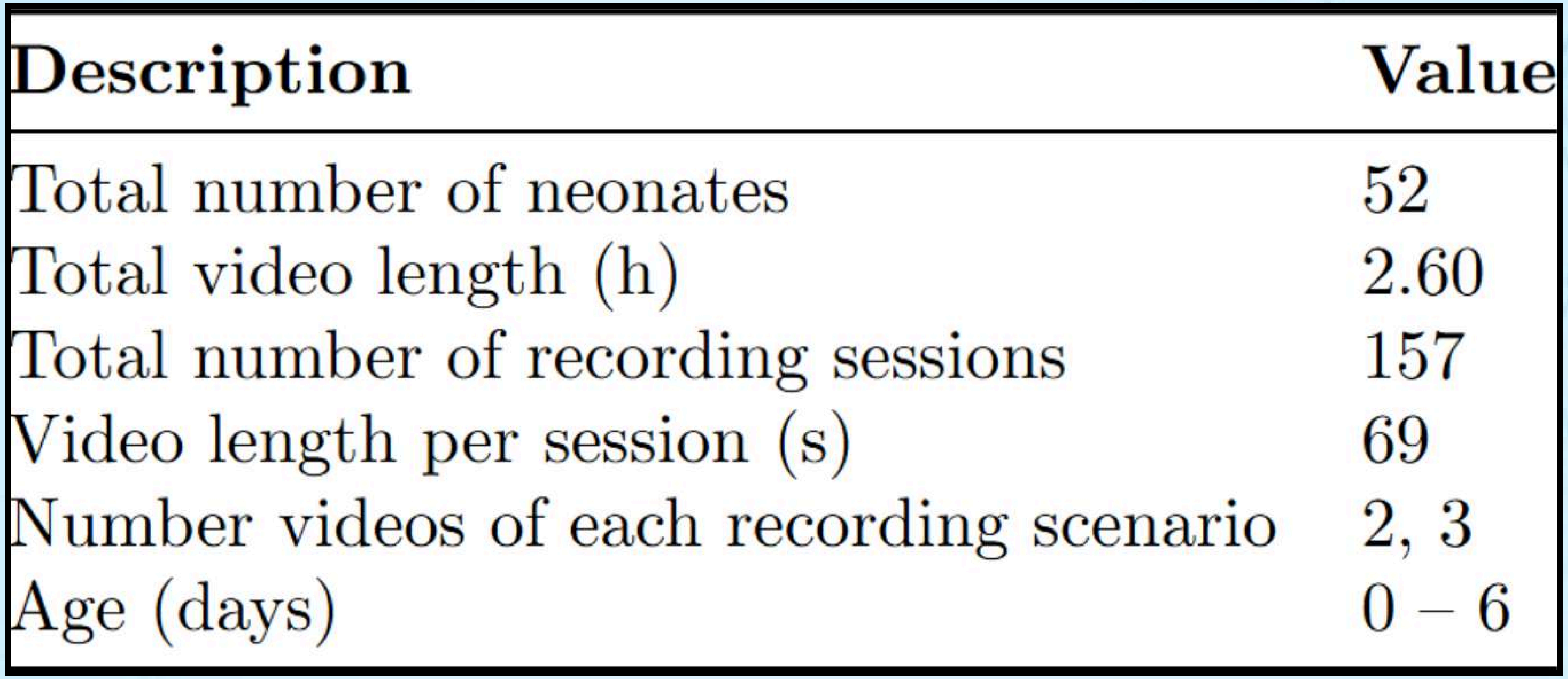
Method	**MAE (bpm↓)**	**SD (bpm↓)**	**MAPE (%↓)**	**TW (s)**
CHROM	6.34	8.02	5.49	12
2SR	5.48	7.69	4.93	6
POS	5.96	10.93	5.12	10
EVM-CNN	4.51	5.78	3.91	8
PRnet	5.66	7.83	4.74	2
NBHRnet-2s	3.97	5.32	3.28	2
NBHRnet-4s	3.83	5.16	3.20	4
NBHRnet-6s	3.76	**5.15**	3.13	6
Ours	**2.97**	5.46	**3.04**	**2**

Comparison of SpO2 estimation methods on NHBR dataset



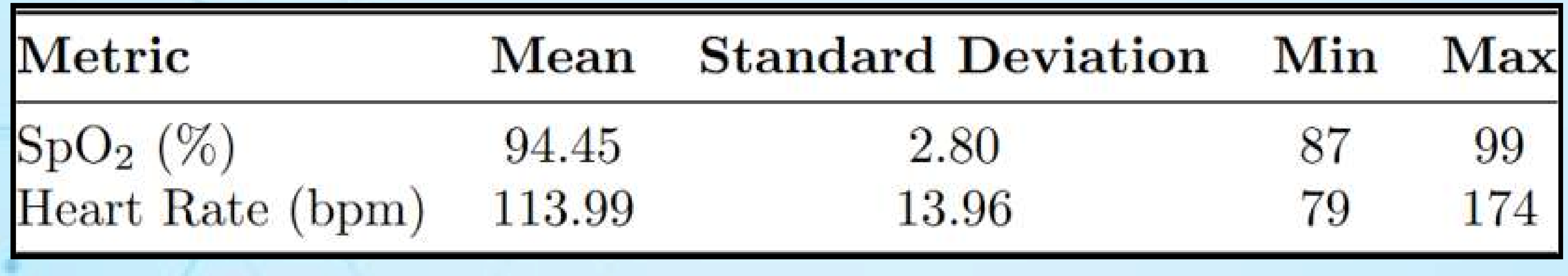
Method	**MAE (%)**	**RMSE (%)**
Past Analytic (Ratio of Ratios)	3.334	5.137
Deep Learning with STMap (EfficientNet-B3 + RGB)	1.274	1.710
Multi-Model Fusion Method	1.000	1.430
Ours	**0.8**	**0.96**

Dataset Details



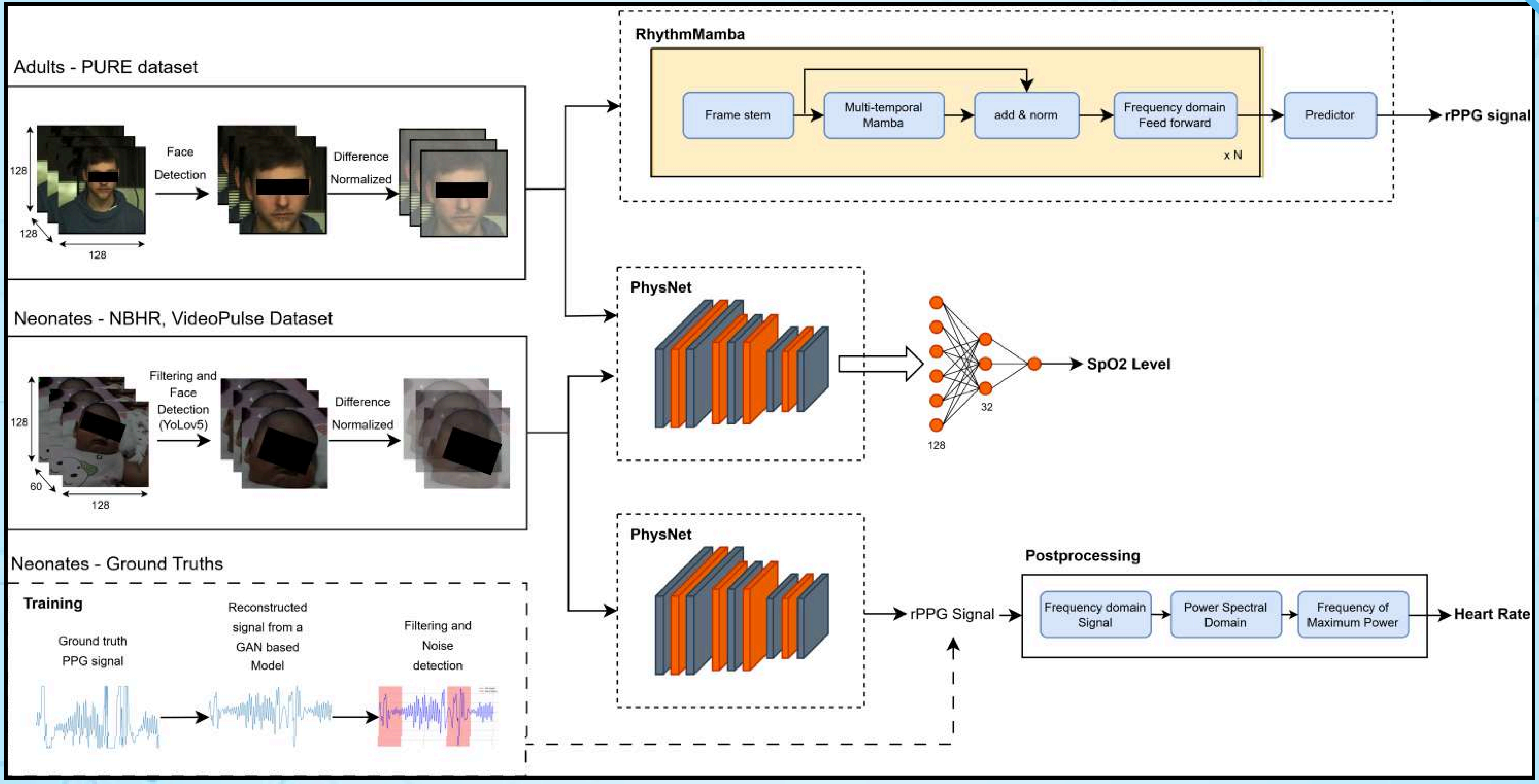
Description	Value
Total number of neonates	52
Total video length (h)	2.60
Total number of recording sessions	157
Video length per session (s)	69
Number videos of each recording scenario	2, 3
Age (days)	0 – 6

Statistics of the Dataset

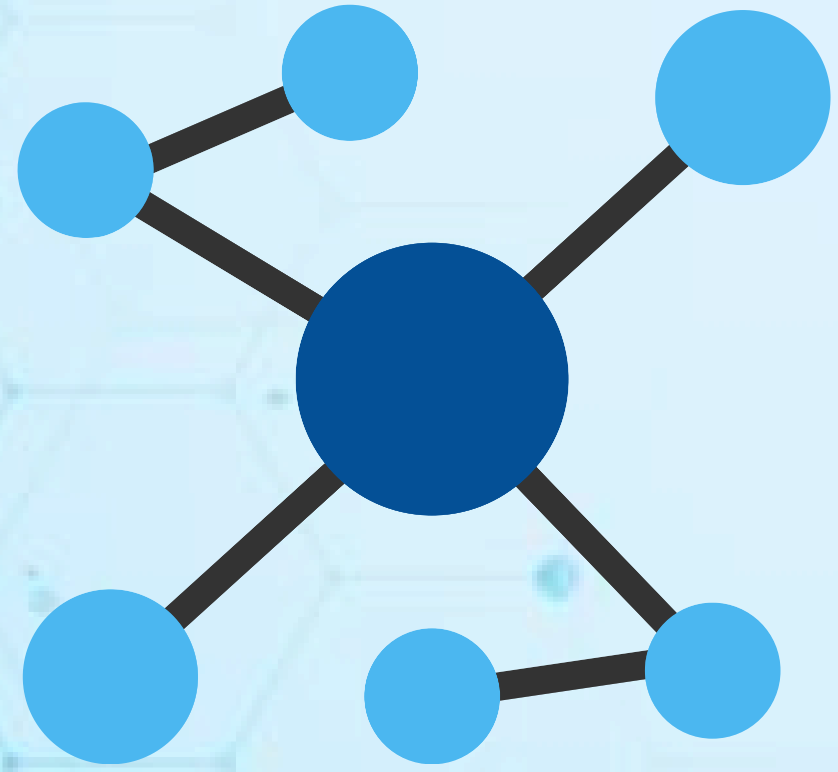


Metric	Mean	Standard Deviation	Min	Max
SpO_2 (%)	94.45	2.80	87	99
Heart Rate (bpm)	113.99	13.96	79	174

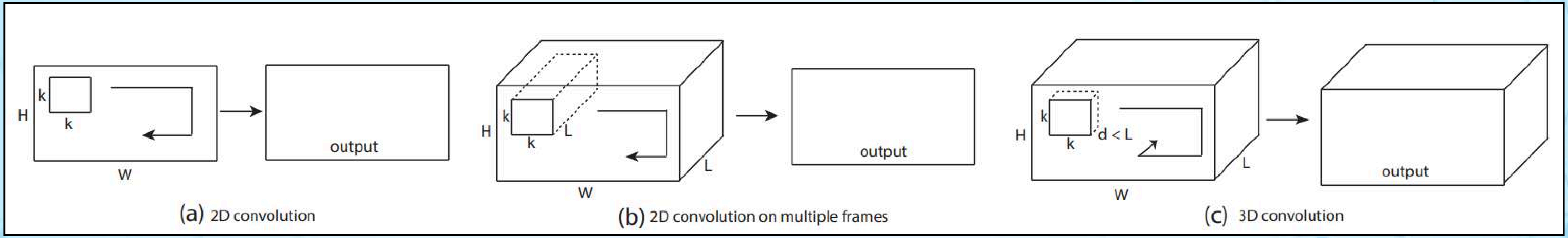
Model Pipeline of Previous Group



ARCHITECTURES



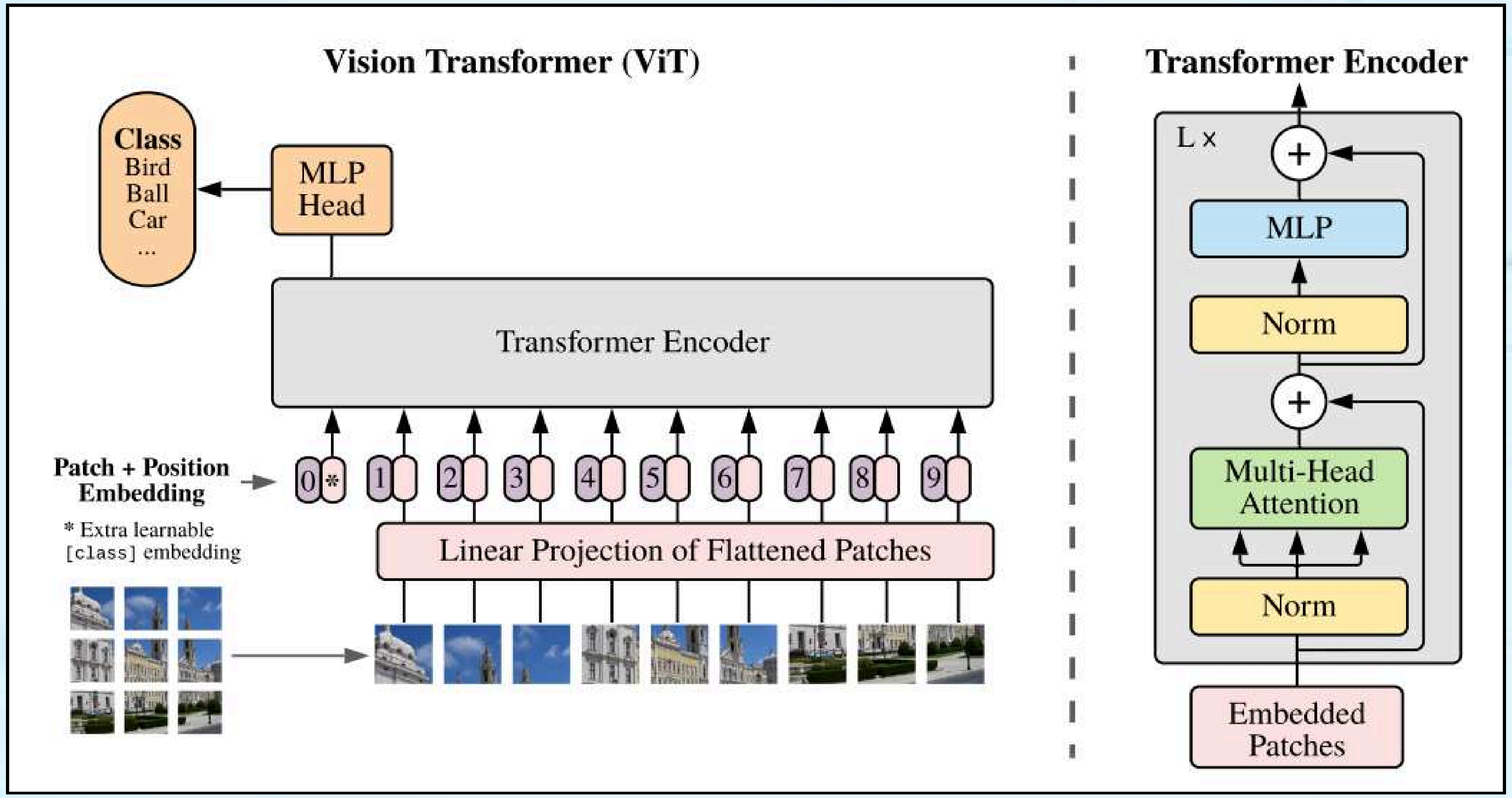
3D CNN



- **Extends 2D CNNs by adding a third dimension—time or depth—to capture spatiotemporal features from volumetric or sequential data such as videos or medical scans.**

D. Tran, L. Bourdev, R. Fergus, L. Torresani and M. Paluri, "Learning Spatiotemporal Features with 3D Convolutional Networks," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 4489-4497, doi: 10.1109/ICCV.2015.510.

Vision Transformer (ViT)

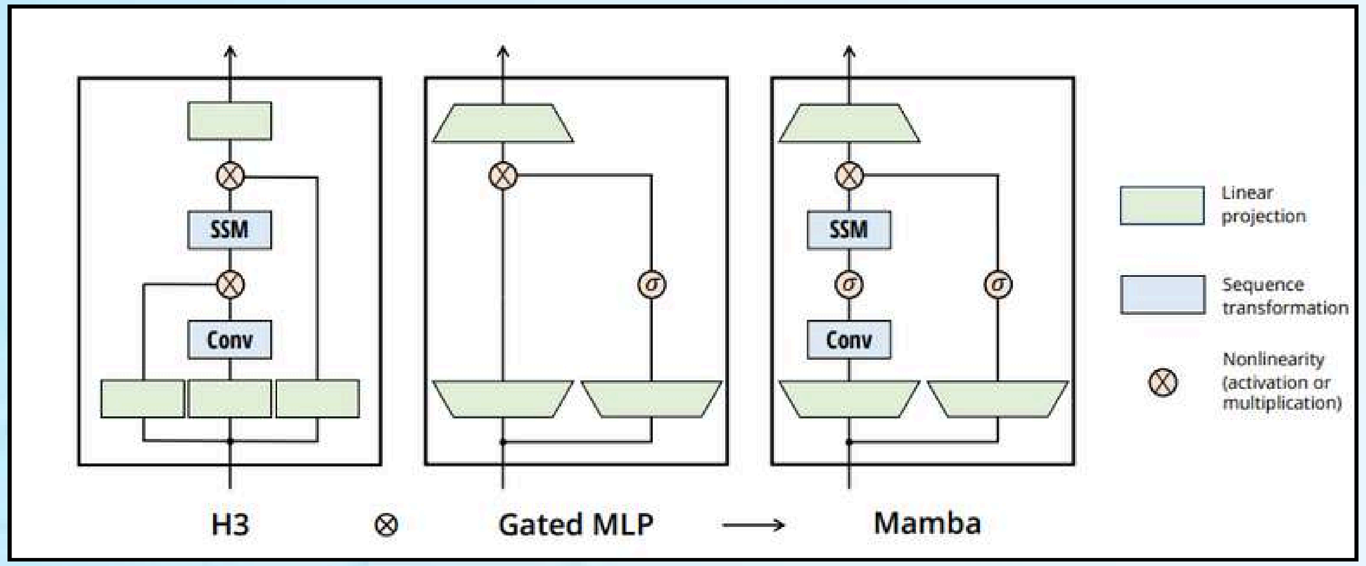


- **Applies the Transformer architecture (originally for NLP) to image classification by treating images as sequences of patches, not pixels.**

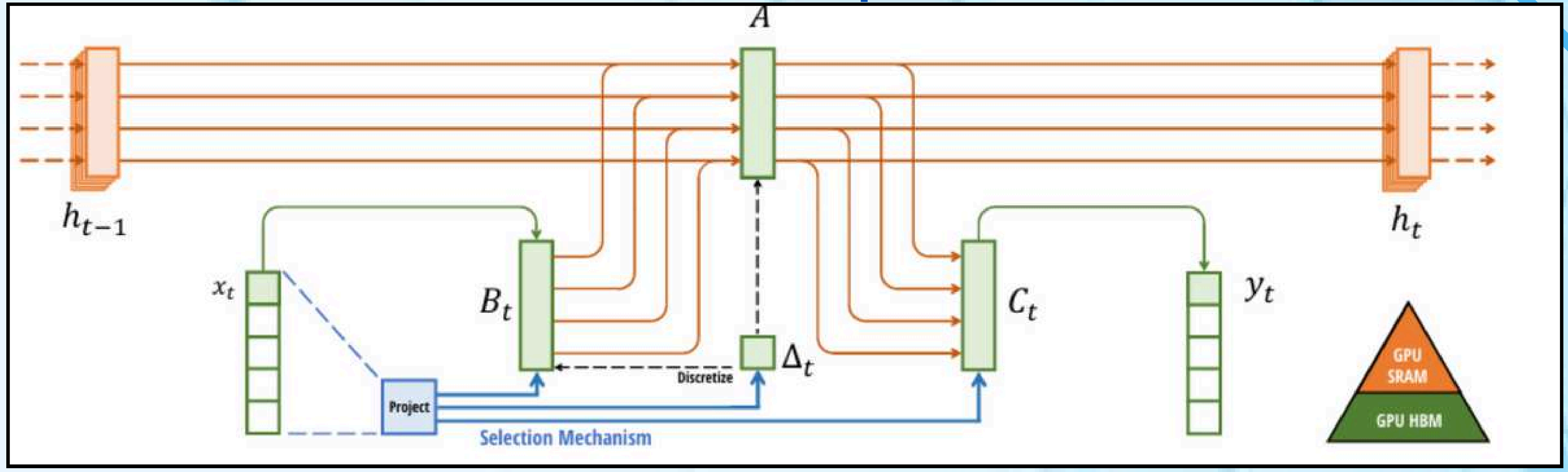
A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," International Conference on Learning Representations, 2021. [Online]. Available: https://openreview.net/forum?id=YicbFdNTTy

Mamba

Architecture



Selective State Space(SSM)



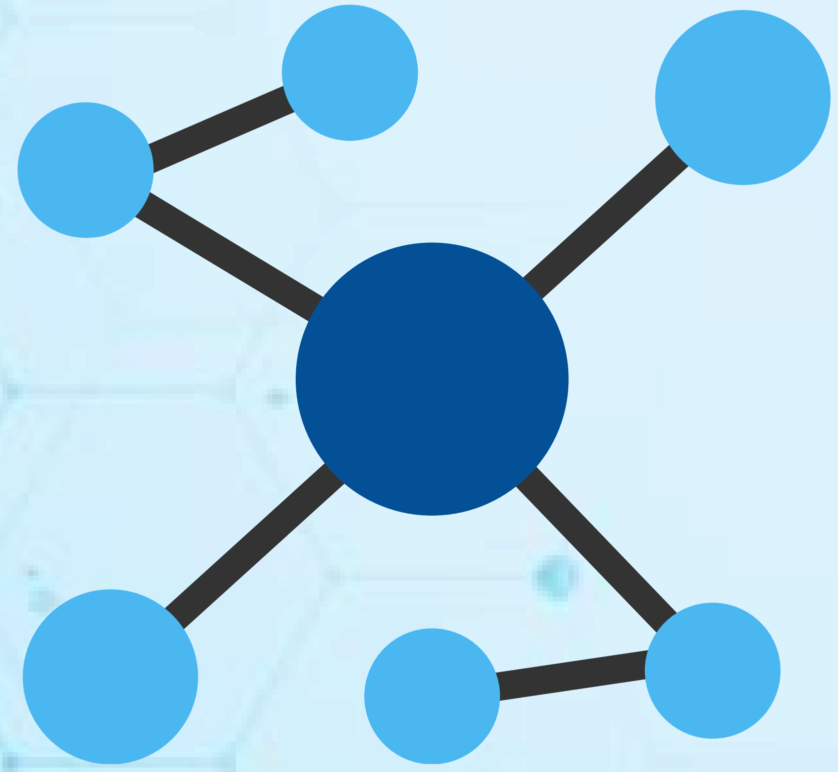
- **A state space model (SSM) designed for efficient long-sequence modeling, offering linear time computation and strong performance on sequential data, without using attention.**

$$
\begin{aligned}
h_t &= A h_{t-1} + B x_t \\
y_t &= C h_t
\end{aligned}
$$

$$
\begin{aligned}
K &= (CB, CAB, \ldots, CA^kB, \ldots) \\
y &= x * K
\end{aligned}
$$

A. Gu and T. Dao, “Mamba: Linear-time sequence modeling with selective state spaces,” 2024. [Online]. Available: https://openreview.net/forum?id=AL1fq05o7H

ROI SELECTION



ROI Selection

rPPG

- The **forehead and cheek** regions are ideal for rPPG measurement as they have richer blood volume and offer more reliable signals. These areas are also less affected by facial muscle movements from expressions or talking

Jaundice Status

- Jaundice in neonates first appears on the face, with the **forehead** showing early signs of yellowing as bilirubin levels rise

WimVerkruysse, Lars O Svaasand, and J Stuart Nelson. Re mote plethysmographic imaging using ambient light. Optics express, 2008.
Khanam F-T-Z, Al-Naji A, Perera AG, Wang D, Chahl J. Non-invasive and non-contact automatic jaundice detection of infants based on random forest. Comput Methods Biomech Biomed Eng: Imaging Vis 2023;11(6):2516–29. http: //dx.doi.org/10.1080/21681163.2023.2244601.

AI MODEL INFERENCE

Model Running Time

Total Prediction Time

↓

face detection + segmentation + model inference

Factors Affecting Face Detection Inference Time

1. Model Size and Complexity
    - YOLOv5n: Lightweight, faster (~20–40 ms/frame with INT8).
    - YOLOv5s: Larger, more accurate, but slower (~60–120 ms/frame with INT8)

2. Input Video Resolution
    - Higher resolutions (e.g., 1080p or 4K): Increase inference time significantly.
    - Optimal input for YOLOv5: 640×640 or 416×416

3. Inference Backend & Hardware Acceleration
    - TFLite + NNAPI (INT8):
        - Uses Android's Neural Networks API for acceleration.
        - Leverages the S25 Ultra's NPU for faster, efficient inference.
    - ONNX + SNPE (INT8):
        - Uses Qualcomm Snapdragon's Neural Processing Engine.

Factors Affecting Segmentation Inference Time

1. Segmentation Model Type and Accuracy
   - Lightweight Models (e.g., DeepLabv3-Mobilenet):
     - Faster but less accurate ROI boundaries.
   - Heavyweight Models (e.g., Mask R-CNN, DeepLabv3+ ResNet101):
     - More precise ROI segmentation, but slower inference.

2. Inference Backend & Hardware Optimization
   - TFLite + NNAPI (INT8):
     - Uses Android's Neural Networks API.
     - Leverages the Galaxy S25 Ultra's NPU for low-latency segmentation and fast ROI selection.
   - ONNX + SNPE (INT8):
     - Uses Qualcomm Snapdragon's Neural Processing Engine for optimized mobile inference.

Face Detection & Segmentation on Mobile

To run our models on-device, we can use lightweight frameworks such as:

- TensorFlow Lite, which supports Android and iOS,
- ONNX Runtime, which provides cross platform compatibility,
- CoreML for iOS deployment and MediaPipe, which offers fast, prebuilt pipelines for face tracking and segmentation.

Why S25 Ultra ?

- Equipped with a dedicated NPU and flagship Snapdragon processor.
- Offers 12GB RAM and hardware-accelerated AI support.
- Unlike most phones, the S25 Ultra can handle simultaneous video capture and model execution (face detection, segmentation, signal extraction) without lag.
- Outperforms competitors like iPhone 16 and Pixel 9 Pro in memory size, NPU capabilities, and frame-by-frame video processing for health applications.

Multi Threading

- We plan to run our two preprocessing pipelines simultaneously.
- We can use multi threading for that.
- This enhances performance and responsiveness of our mobile application.

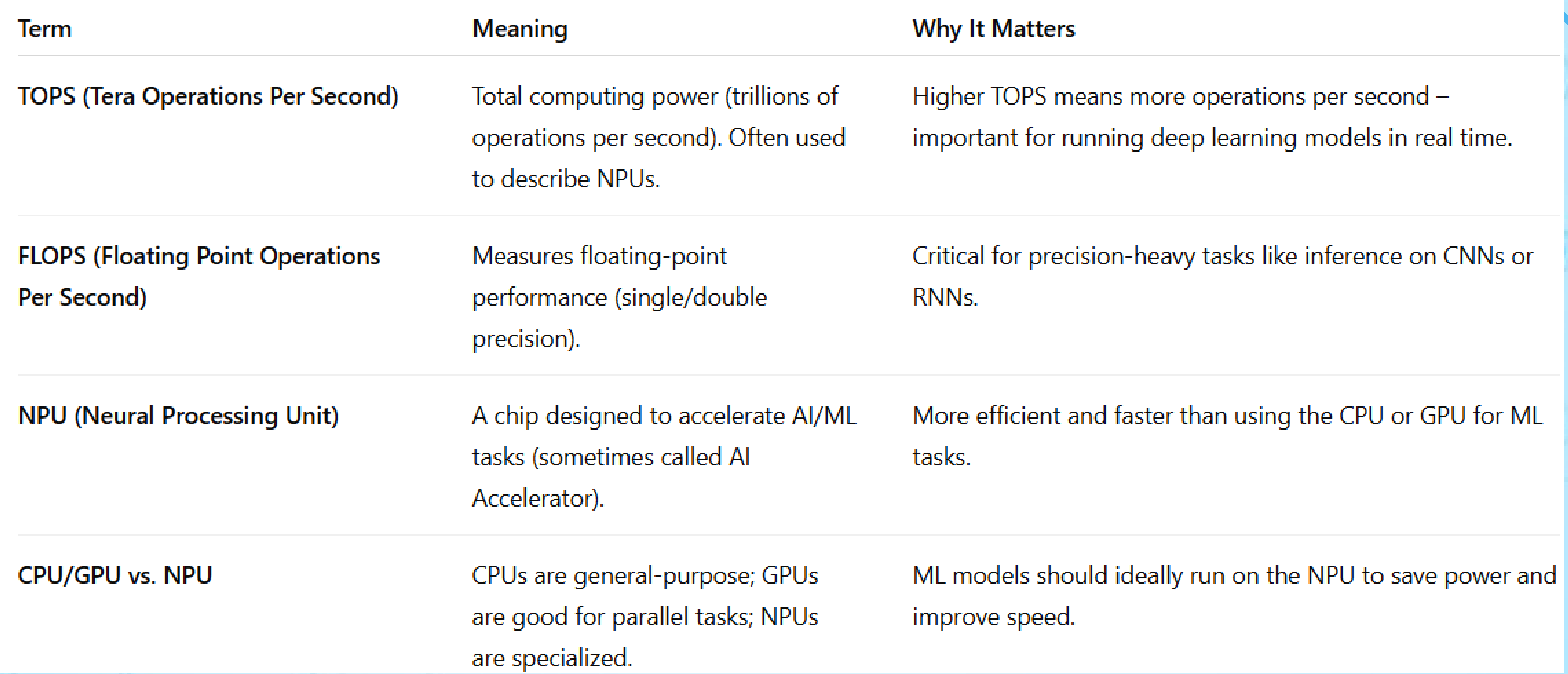
A study on performance optimization of multi-threading and concurrency handling techniques in android applications

Mei Liu[*] and Qun Wang

Shandong Xiehe University, 250100 JiNan, Shandong, China

A study on performance optimization of multi-threading and concurrency handling techniques in android applications Mei Liu* and Qun Wang , MATEC Web of Conferences 395, 01042 (2024) https://doi.org/10.1051/matecconf/202439501042

Factors to consider when running ML models in mobile phone



Term	Meaning	Why It Matters
TOPS (Tera Operations Per Second)	Total computing power (trillions of operations per second). Often used to describe NPUs.	Higher TOPS means more operations per second – important for running deep learning models in real time.
FLOPS (Floating Point Operations Per Second)	Measures floating-point performance (single/double precision).	Critical for precision-heavy tasks like inference on CNNs or RNNs.
NPU (Neural Processing Unit)	A chip designed to accelerate AI/ML tasks (sometimes called AI Accelerator).	More efficient and faster than using the CPU or GPU for ML tasks.
CPU/GPU vs. NPU	CPUs are general-purpose; GPUs are good for parallel tasks; NPUs are specialized.	ML models should ideally run on the NPU to save power and improve speed.